

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

MÁRCIO SYGEAKS FRAZÃO BARROS

**CLASSIFICAÇÃO DE FALHAS GEOLÓGICA EM DADOS SÍSMICO
USANDO COVARIÓGRAMA E APRENDIZADO DE MÁQUINA**

São Luís
2013

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

MÁRCIO SYGEEKS FRAZÃO BARROS

**CLASSIFICAÇÃO DE FALHAS GEOLÓGICA EM DADOS SÍSMICO
USANDO COVARIOGRAMA E APRENDIZADO DE MÁQUINA**

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Anselmo Cardoso Paiva

São Luís

2013

Barros, Márcio Sygeaks Frazão

Classificação de falhas geológicas em dados sísmicos usando covariograma e aprendizado de máquina / Márcio Sygeaks Frazão Barros. – 2013.

53f.

Orientador: Anselmo Cardoso Paiva.

Monografia (Graduação em Ciência da Computação) – Universidade Federal do Maranhão, 2013.

1.Dados sísmico 2.Falhas geológicas 3.Aprendizado de máquina 4.SVM
5.Função geoestatística 6.Covariograma I. Título.

CDU 550.832.4

Márcio Sygeaks Frazão Barros

**Classificação de Falhas Geológica em Dados Sísmico usando
Covariograma e Aprendizado de Máquina**

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Aprovada em: 18 / 02 / 2013

BANCA EXAMINADORA



Orientador: Anselmo Cardoso Paiva

Prof. Dr. em Informática – UFMA

Universidade Federal do Maranhão



Aristófanes Córrea Silva

Prof. Dr. em Informática – UFMA

Universidade Federal do Maranhão



Geraldo Braz Junior

Prof. em Informática – UFMA

Universidade Federal do Maranhão

AGRADECIMENTOS

Em primeiro lugar o meu Deus por ter me dado forças e iluminando meu caminho para que pudesse concluir mais uma etapa da minha vida;

Ao meu pai, por todo amor e dedicação que sempre teve comigo, homem pelo qual tenho maior orgulho de chamar de pai, meu eterno agradecimento pelos momentos em que estive ao meu lado, me apoiando e me fazendo acreditar que nada é impossível, pessoa que sigo como exemplo, pai dedicado, amigo, batalhador;

A minha mãe, por ser tão dedicada e amiga, por ser a pessoa que mais me apóia e acredita na minha capacidade, meu agradecimento pelas horas em que ficou ao meu lado não me deixando desistir e me mostrando que sou capaz de chegar onde desejo, sem dúvida foi quem me deu o maior incentivo para conseguir concluir esse trabalho;

Aos meus irmãos pelo carinho e atenção que sempre tiveram comigo, em especial ao meu irmão Ivanildo Barros que sempre me apoiando em todos os momentos, enfim por todos os conselhos e pela confiança em mim depositada meu imenso agradecimento;

Aos professores, orientador Dr. Anselmo Cardoso Paiva, Dr. Aristófanés Corrêa Silva, Geraldo Braz Junior e João Dallyson Sousa de Almeida pela orientação, dedicação, amizade, incentivo, conselhos e apoio durante estes anos de convivência e trabalho;

Aos amigos que fiz durante o curso, pela verdadeira amizade que construímos em particular aqueles que estavam sempre ao meu lado (Antônio Gomes, Afonso Lopes, Pablo Fernandes, Jefferson Sá, Valéria Priscilla, Wellson, Nadson Timbó e Wagner) por todos os momentos que passamos durante esses cinco anos meu especial agradecimento. Sem vocês essa trajetória não seria tão prazerosa;

A todos os professores do curso Ciência da Computação, pela paciência, dedicação e ensinamentos disponibilizados nas aulas, cada um de forma especial contribuiu para a conclusão desse trabalho e conseqüentemente para minha formação profissional;

Ao Programa de Recursos Humanos da ANP – PRH 39 pela bolsa de iniciação científica concedida.

Por fim, gostaria de agradecer a todos aos meus amigos e familiares da cidade de Monção, pelo carinho e pela compreensão nos momentos em que a dedicação aos estudos foi exclusiva, a todos que contribuíram direta ou indiretamente para que esse trabalho fosse realizado, meu eterno AGRATECIMENTO.

RESUMO

O petróleo é a principal fonte de energia utilizada no mundo, mas também exige um custo bastante elevado para sua obtenção. A perfuração de um poço de petróleo pode atingir, em uma estimativa, até 85% do custo total da exploração. Portanto, é necessário ter o máximo de informações possíveis das condições geológicas da área antes de iniciar a perfuração de um poço. Uma das etapas de obtenção dessa fonte energética é a prospecção que inicia pelos estudos geológicos de observação até a obtenção de dados sísmicos e a perfuração do poço pioneiro. A sísmica hoje é uma das ferramentas mais importantes na prospecção do petróleo, mas os dados sísmicos coletados necessitam de tratamentos através de programas computacionais para poderem ser interpretados pelos geólogos e geofísicos. O objetivo desse trabalho é desenvolver um método usando a função Covariograma como descritores de característica e aprendizado de máquina, que auxilie o intérprete na geração de um modelo que melhor representa a estrutura geológica para classificação de falhas em dados sísmicos. O método obteve resultados de sensibilidade igual a 91.10%, especificidade igual a 93.92% e acurácia igual a 92.23%

Palavras-chave: Dados sísmicos. Falhas geológicas. Aprendizado de Máquina. SVM. Função geoestatística.

ABSTRACT

Oil is the main source of energy used worldwide, but also requires a high cost to obtain. Drilling an oil well can achieve in an estimate of 85% of the total cost of operation. Therefore, you need as much information as possible on geological conditions of the area before starting the drilling of a well. One of the steps to obtain this energy source is the prospect that starts by geological observation to obtain seismic data and the drilling of wildcat well. The seismic is today one of the most important tools in the exploration of oil, but the seismic data collected require treatment by computer programs so that they can be interpreted by geologists and geophysicists. The aim of this work is to develop a method using the function as Covariograma feature descriptors and machine learning, which helps the performer to generate a model that best represents the geological structure for fault detection in seismic data. The method results obtained sensitivity equal to 91.10%, specificity of 93.92% and accuracy equal to 92.23%.

Keywords: Seismic data. geological faults. Machine Learning. SVM. Function geostatistics.

Sumário

1. INTRODUÇÃO.....	12
2. FUNDAMENTAÇÃO TEÓRICA	14
2.1. Dados Sísmicos	14
2.1.1. Sísmica de Reflexão	15
2.1.2. Métodos Sísmicos.....	18
2.1.3 Aquisição	19
2.1.4. Processamento.....	22
2.1.5. Interpretação.....	24
2.2. Atributo Sísmico	25
2.2.1 Importância do atributo sísmico na geofísica	26
2.2.2. Similaridade	27
2.2.3. Energia	28
2.2.4. Funções Geoestatísticas	28
2.2.4.1. Covariograma	30
2.3. Reconhecimento de Padrões	31
2.3.1. Máquina de Vetor de Suporte	32
2.3.2. Considerações sobre a Escolha do Classificador.....	33
2.3.3. Validação de Resultados	35
3. MATERIAIS E MÉTODOS	37
3.1. Aquisição e tratamento das imagens sísmicas	37
3.2. Cálculo dos atributos sísmico de falha	37
3.2.1 Cálculo da Similaridade Vertical (SV).....	38
3.2.2 Cálculo do Atributo Energia Vertical (EV).....	39
3.2.3. Cálculo do atributo Covariograma.	41
3.2.3.1. Extração de Característica	41
3.2.3.2. Vetor de Características.....	42
3.4. Classificação usando MVS.	43
4. RESULTADOS	46
5. CONCLUSÃO.....	49
REFERÊNCIAS.....	51

Lista de Figuras

Figura 2.1 Estrutura capeadora[w. Teixeira ET AL. 2000].	15
Figura 2.2 – Etapas de exploração baseada na sísmica de reflexão [Silva 2004].	19
Figura 2.3: Aquisição Sísmica Terrestre.	20
Figura 2.4: Aquisição Sísmica Marítima (FLTC SERVICES LTD., 2012).	20
Figura 2.5: Aquisição Marítima [Oilfield].	22
Figura 2.6: Traço sísmico (esquerda), linha sísmica (centro) e volume sísmico (direita). [Kristofer M. Tingdahl & Matthijs de Rooij]	23
Figura 2.7 – Modelo geológico. Adaptada de [Robison, Treitel 1980].	24
Figura 2.8: Atributos utilizados por Tingdahl [Tingdahl & de Rooij, 2005].	27
Figura 2.9: Definição do vetor distância.	30
Figura 2.10: Parâmetros utilizados para cálculo das função geoestatísticas.[Aristófanés 2004].	31
Figura 2.11: Hiperplano ótimo.[Lima, 2004]	34
Figura 3.1: Fatia gerada a partir de dados da amplitude.	38
Figura 3.2: Cálculo da Similaridade Vertical.	39
Figura 3.3: Similaridade Vertical	39
Figura 3.4: Cálculo da Energia Vertical.	40
Figura 3.5: Energia Vertical	40
Figura 3.6: Distância e direções dos lags. [CAMARGO 1998].	42
Figura 3.7: Fluxo de atividade realizado durante a etapa de classificação desde a base de características originais, estimação de parâmetros e reconhecimento de padrão.	43
Figura 3.8: Resultados obtidos nos testes.	48

Lista de Tabelas

Tabela 4.1: Resultados da acurácia para 2500 amostra calculada para cada atributos.	46
Tabela 4.2: Resultados da acurácia para 2500 amostra calculada para associação dos atributos.	47
Tabela 4.3: Melhor resultados obtidos para a classificação.....	47

1. INTRODUÇÃO

A descoberta de uma jazida de petróleo em uma nova área é uma tarefa que envolve um longo e dispendioso estudo e análise de dados geofísicos e geológicos das bacias sedimentares.

Os geólogos e geofísicos são os profissionais que decidem propôr a perfuração de um poço, que é a etapa que mais investimentos exigem em todo o processo de prospecção.

Um programa de prospecção visa fundamentalmente dois objetivos: (1) localizar dentro de uma bacia sedimentar as situações geológicas que tenham condições para a acumulação de petróleo; e (2) verificar qual, dentre estas situações, possui mais chance de conter petróleo.

A identificação de uma área favorável à acumulação de petróleo é realizada através de métodos geológicos e geofísicos, que, atuando em conjunto, conseguem indicar o local mais propício para a perfuração. O programa desenvolvido durante a fase de prospecção fornece uma quantidade muito grande de informações técnicas.

A primeira etapa de um programa exploratório é a realização de um estudo geológico com o propósito de reconstituir as condições de formação e acumulação de hidrocarbonetos em uma determinada região.

Através do mapeamento das rochas que afloram na superfície, é possível reconhecer e delimitar as bacias sedimentares e identificar algumas estruturas capazes de acumular hidrocarbonetos. As informações geológicas e geofísicas obtidas a partir de poço exploratórios são de enorme importância para a prospecção, pois permitem reconhecer as rochas que não afloram na superfície e aferir e calibrar os processos indiretos de pesquisas como os métodos sísmicos.

O método sísmico de reflexão é o método de prospecção mais utilizado atualmente na indústria de petróleo, pois fornece alta definição das feições geológicas em subsuperfície propícias à acumulação de hidrocarbonetos, mais de 90% dos investimentos em prospecção são aplicados em sísmica de reflexão.

O objetivo desse trabalho é desenvolver um método usando a função Covariograma como descritores de característica e aprendizado de máquina, que auxilie o intérprete na geração de um modelo que melhor representa a estrutura geológica para classificação de falhas em dados sísmicos. Dessa forma, este

trabalho está estruturado em 5 Capítulos. O Capítulo 2 apresenta a Fundamentação teórica. O Capítulo 3, a metodologia utilizada para se chegar à solução, indo desde a seleção e preparação dos dados, cálculo do método geoestatístico covariograma como descritores de características, e também cálculo do atributo de Energia vertical e Similaridade Vertical e por fim a classificação. O Capítulo 4, os resultados apresentados. No Capítulo 5 fazemos as considerações finais do trabalho realizado.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo será dada uma visão geral de alguns conceitos que serão de fundamental importância na compreensão dos capítulos seguintes. Aborda-se o conceito de dados sísmico, sísmica de reflexão, aquisição, processamento de imagem, atributos sísmico, método de análise de textura baseado em medidas geoestatísticas: covariograma e método de classificação e reconhecimento de padrão utilizando Máquinas de Vetores de Suporte.

2.1. Dados Sísmicos

A sísmica de reflexão utilizada pelos geofísicos tem como objetivo a descoberta de reservatórios de hidrocarbonetos. Segundo Gehardt [Gerhardt, 1998], a grande maioria dos acúmulos de petróleo (hidrocarbonetos como óleo e gás) se encontra em rochas sedimentares.

A existência de acumulações de petróleo depende das características e do arranjo de certos tipos de rochas sedimentares no subsolo. Basicamente, é preciso que existam rochas geradoras que contenham a matéria-prima que se transforma em petróleo e rochas-reservatório, ou seja, aquelas que possuem espaços vazios, chamados poros, capazes de armazenar o petróleo. Essas rochas são envolvidas em armadilhas chamadas trapas, compartilhamentos isolados no subsolo onde não tem condições de escapar.

A ausência de qualquer um desses elementos impossibilita a existência de uma acumulação petrolífera. Logo, a existência de uma bacia sedimentar não garante, por si só, a presença de jazidas de petróleo.

As rochas geradoras são assim chamadas por tratar-se de um mineral formado principalmente pelo acúmulo de fragmentos de outros minerais e detritos orgânicos, e que, quando se encontra num ambiente de pouca permeabilidade o que inibe a ação de água circulante e diminui a quantidade de oxigênio existente cria as condições necessárias para a formação do petróleo.

Após o processo de formação do petróleo, para que o mesmo se acumule, formando posteriormente um reservatório, é necessário que após a geração ocorra a migração do petróleo, e que no percurso desta migração exista alguma armadilha geológica que permita a acumulação do óleo. A Figura 2.1 apresenta alguns tipos de reservatórios estrutural, estratigráfica e combinada.

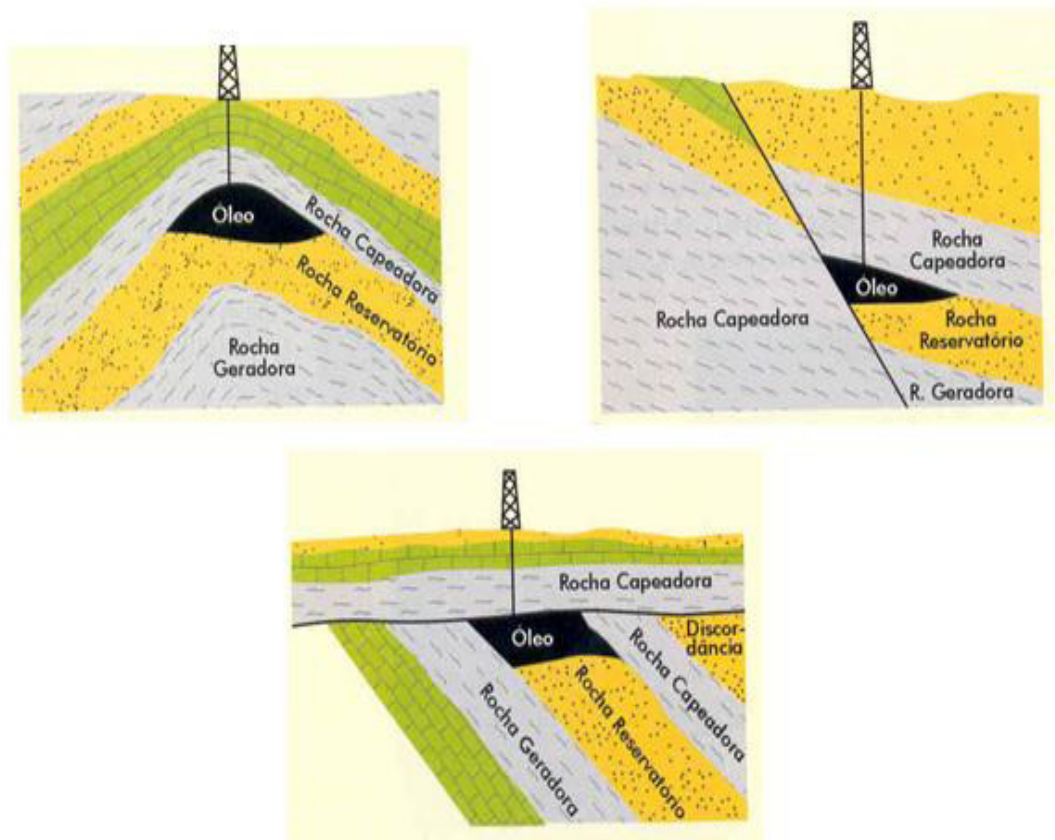


Figura 2.1 Estrutura capeadora[w. Teixeira ET AL. 2000].

O petróleo migra da rocha geradora para outra rocha, porosa e permeável, chamada Rocha Reservatório, e continua seu fluxo no interior da mesma, até ser contido por uma armadilha, isto é, uma estrutura geológica compreendida dentro de uma rocha selante (impermeável), que permita que o óleo ali se confine.

2.1.1. Sísmica de Reflexão

Sísmica de reflexão é um método de prospecção geofísica que utiliza os princípios da sismologia para estimar as propriedades da sub-superfície da Terra com base na reflexão de ondas sísmicas. Este método requer a utilização

de uma fonte sísmica de energia controlada, como por exemplo, um explosivo. Ao determinar o tempo que uma onda refletida demora até atingir um receptor, é possível estimar a profundidade da estrutura que gerou a reflexão.

Trata-se de um método utilizado extensivamente na prospecção de hidrocarbonetos (petróleo e gás natural) e de outros recursos minerais como carvão, minérios metálicos e energia geotérmica. É ainda utilizado em estudos básicos sobre a natureza e origem das rochas que compõem a crosta terrestre.

O método sísmico de reflexão é o método de exploração mais utilizado atualmente na indústria do petróleo, destaca-se pelo alto grau de eficiência, a um custo relativamente baixo. Mais de 90% dos investimentos em prospecção são aplicados em sísmica de reflexão. Por este método obtém-se excelente definição da formação geológica da sub-superfície, permitindo a análise da probabilidade do acúmulo de hidrocarbonetos.

O levantamento sísmico iniciado com a geração de ondas elásticas, através de fontes artificiais, que se propagam pelo interior da Terra, onde são refletidas e refratadas nas interfaces que separam rochas de diferentes constituições petrofísicas, e retornam à superfície. As reflexões são captadas por equipamentos especiais denominados geofones (para registros em terra) ou hidrofones (para registros no mar), os quais convertem as vibrações mecânicas em oscilações elétricas que são transmitidas e registradas nos sismógrafos.

2.1.1.1. Tipos de Ondas Sísmicas e Velocidade de Propagação

Existem basicamente dois tipos de ondas elásticas. As ondas P, ou compressionais, nas quais o deslocamento do meio se dá na mesma direção de propagação da energia, e as ondas S, ou de cisalhamento, onde o deslocamento do meio é perpendicular à direção de propagação de energia. Até o momento, somente as ondas P vêm sendo utilizadas comercialmente nos levantamentos sísmicos.

A velocidade de propagação das ondas sísmicas é função da densidade e das constantes elásticas do meio. Conseqüentemente depende da constituição mineralógica da rocha, grau de cimentação, estágios de compactação, porosidade,

conteúdo de saturação de fluidos, além de outros fatores como temperatura e presença de microfraturas.

Para todos os fins práticos, a propagação das ondas elásticas é regida pelas mesmas leis da ótica geométrica. Quando uma frente de onda incide sobre uma interface separando duas rochas com velocidades e densidades diferentes, parte da energia incidente é refratada para o meio inferior, e parte da energia se reflete e retorna à superfície. A quantidade de energia que retorna à superfície depende do contraste de impedâncias acústicas (produto da densidade pela velocidade) dos dois meios e do ângulo de incidência.

2.1.1.2. Fontes e Receptores Sísmicos

As fontes de energia sísmica mais utilizadas são a dinamite e o vibrador, em terra, e canhões de ar comprimido, em levantamentos marítimos. Cada uma destas fontes emite um pulso característico conhecido como assinatura da fonte que se propaga em todas as direções. Estes pulsos elásticos ou detonações são de duração ou comprimento muito pequeno, da ordem de 200 milissegundos, e se refletem e propagam em cada uma das camadas geológicas em profundidade, retornando à superfície com informações valiosas para a pesquisa de petróleo.

Os receptores utilizados para registrar as reflexões destes pulsos são basicamente de dois tipos: eletromagnéticos (geofones) para registros em terra, e de pressão (hidrofones) para levantamentos na água. O primeiro é composto por uma bobina suspensa dentro de um campo magnético gerado por um potente ímã acondicionado em invólucro impermeável, que é firmemente cravado à superfície da terra. Quando uma onda sísmica atinge o geofone, o movimento relativo entre a bobina e o ímã gera uma corrente elétrica induzida que é proporcional a vários fatores, inclusive à amplitude da onda incidente.

Os hidrofones utilizam cristais piezoelétricos, que geram uma corrente elétrica proporcional à variação de pressão produzida pelas ondas acústicas na água. Estes receptores, a exemplo dos geofones, devem reproduzir o mais fielmente possível as vibrações mecânicas na forma de oscilações elétricas. Essas oscilações elétricas são transmitidas até o sismógrafo, onde são digitalizadas, multiplexadas e

registradas (ou retransmitidas via satélite para uma central de computadores) após severo deperamento e amplificação eletrônicos.

2.1.2. Métodos Sísmicos

O método sísmico tem como objetivo principal a formação de modelos de dados que, depois de processados e organizados, retenham informações relevantes a respeito da geologia da região onde foram adquiridos. Dentre os modelos existentes, O método direto de identificação de reservatórios de hidrocarbonetos em superfícies consiste na perfuração de poços, chamados poços exploratórios. Embora, obviamente, a perfuração seja necessária para a extração do óleo e gás, ela nos fornece dados pontuais e envolve um elevado custo. Os métodos indiretos de investigação são mais baratos e geram informações sobre grandes áreas ou volumes da sub-superfície. Dentre estes, o mais empregado é a sísmica de reflexão. Atualmente, a maioria dos levantamentos sísmicos é tridimensional, isto é, coletam sinais referentes a porções volumétricas da sub-superfície. Os dados adquiridos são submetidos ao chamado processamento sísmico, que compreende um conjunto de técnicas que atuam sobre os dados de forma a retirar ou minimizar eventos indesejáveis, aumentando a relação sinal-ruído. O objetivo do processamento sísmico é facilitar a interpretação dos dados sísmicos.

A interpretação sísmica começa pelo mapeamento da estrutura em larga escala dos dados. Essa interpretação estrutural consiste principalmente em criar horizontes e determinar superfícies de falha. Horizontes são superfícies criadas pelo intérprete através da seleção de um refletor (interface entre camadas diferentes) e seu acompanhamento ao longo do volume. A fratura de rochas em sub-superfície causada por forças tectônicas é chamada de falha. Falhas geram descontinuidades na estrutura de camadas que tornam a criação de horizontes mais difícil.

As falhas podem ser extremamente importantes na identificação de trapas de origem estrutural. Deslocando uma camada impermeável, uma falha pode criar uma trapa estrutural onde o óleo é aprisionado, ou pode romper a capa selante de uma trapa, gerando um vazamento [Lines et al., 2004]. De forma semelhante, as falhas também são importantes na compreensão do fluxo de fluidos em um

reservatório. Elas podem agir no sentido de gerar um selante, bloqueando o fluxo entre diferentes compartimentos de um reservatório ou podem agir como condutores, permitindo a comunicação entre diferentes compartimentos [Pedersen, 2007].

Segundo Silva [Silva 2004], a exploração de hidrocarbonetos, óleo e gás baseada em sísmica pode ser dividida em três etapas principais: aquisição, processamento e interpretação. A Figura 2.2 ilustra essas etapas.

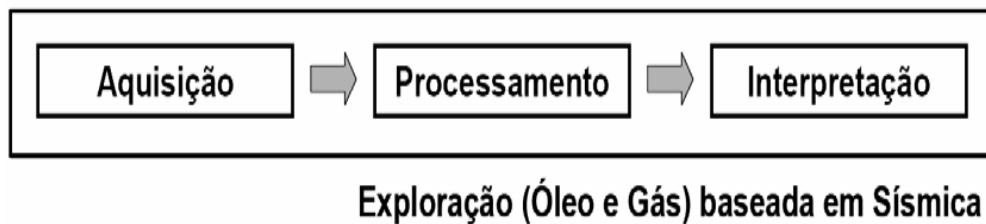


Figura 2.2 – Etapas de exploração baseada na sísmica de reflexão [Silva 2004].

2.1.3. Aquisição

Na etapa de aquisição dos dados são geradas ondas elásticas artificiais (impulsos) de duração relativamente pequena (em torno de 200 milissegundos) em pontos específicos na superfície da área a ser mapeada. Essas perturbações mecânicas são geradas através da utilização de dinamite quando em terra, ou canhões de ar comprimido no caso de regiões marinhas [Moraes, 2007].

Uma perturbação mecânica vibrações elástica ou acústica que se propaga na Terra, quando a onda sísmica alcança camada entre duas rochas com valores de impedância acústica diferente, parte da onde é refratada e continua se propagando para baixo; outra parte é refletida e retorna a superfície [Silva, 2004]. A parte refletida, que retorna á superfície, é captada pelos receptores e tem tanto o tempo de chegada quanto a intensidade medida neste momento, registrado nos sismógrafos armazena as amplitudes das ondas em intervalos de tempo regulares (tipicamente e a 4 milissegundos).

Os receptores podem ser geofones, que medem a velocidade do pulso sísmico que retorna á superfície, na aquisição de dados terrestres, ou hidrofones, que medem a variação de pressão gerada na água [Machado, 2000]. A Figura 2.3 e

Figura 2.4 ilustram um esquema de aquisição sísmica terrestre, e aquisição sísmica marítima.

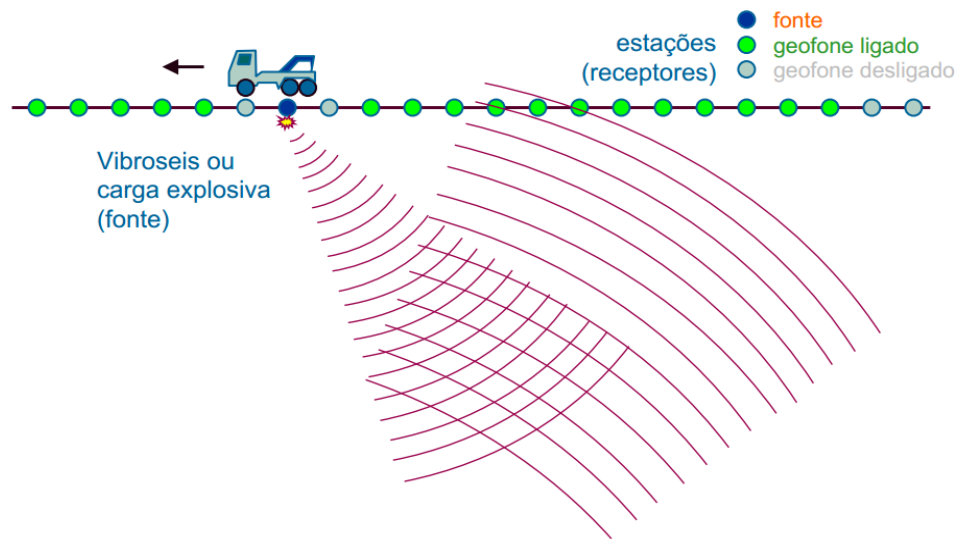


Figura 2.3: Aquisição Sísmica Terrestre.

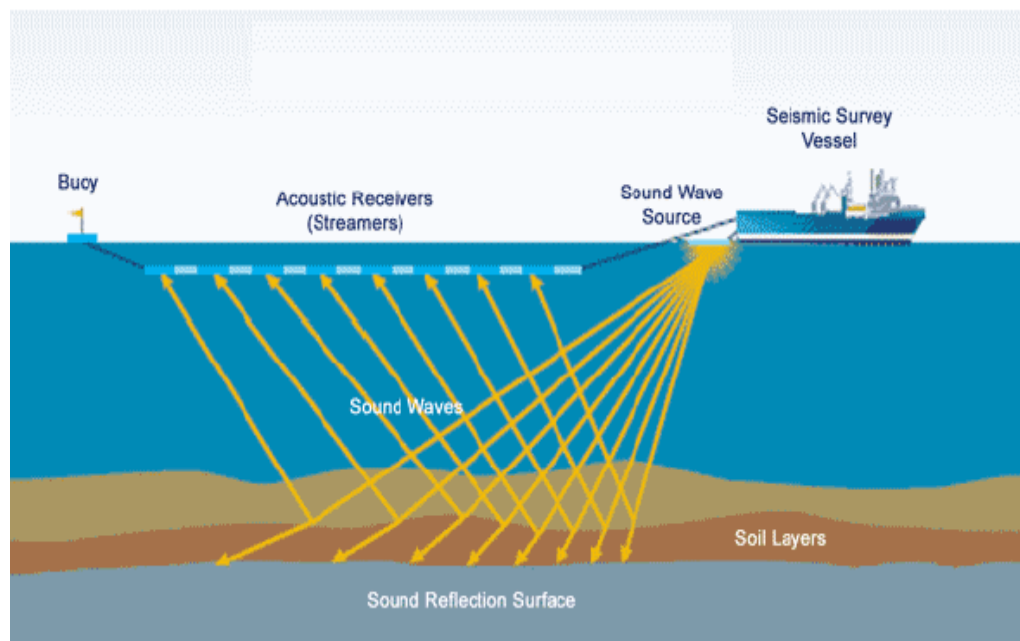


Figura 2.4: Aquisição Sísmica Marítima (FLTC SERVICES LTD., 2012).

A aquisição marítima é realizada a partir de embarcações equipadas com canhões de ar comprimido, rebocando cabos sismógrafos com comprimentos que podem variar entre 4 e 16 km como podemos ver na Figura 2.5. Os cabos sismógrafos são preenchidos com fluídos de baixa densidade para facilitar a flutuação e são equipados com detectores de pressão (hidrofonos), dispostos em

intervalos regulares. Durante o período de aquisição, a área ocupada pelos cabos esticados pode chegar a 10 km².

A aquisição sísmica é iniciada com a geração de energia diretamente na água a partir de disparos do canhão de ar comprimido. A energia que for criada se propaga sob a forma de ondas acústicas, até atingir a profundidade do mar. Ao chegarem ao fundo do mar, essas ondas sofrem um fenômeno físico chamado “partição de energia”, parte da energia é refletida e parte é refratada. A energia refratada é transmitida nas próximas camadas, que conseqüentemente irão sofrer o mesmo fenômeno de partição de energia, sendo que a energia refletida por estas é capturada pelos hidrofones [Chevron, 2007].

A máxima profundidade a ser mapeada no processo de aquisição é determinada pelo tempo de registro das reflexões pelos sensores e pela velocidade de propagação dos sinais sísmicos através das camadas de rocha. Para aquisição terrestre, esse tempo de amostragem é tipicamente de 4 segundos a partir do instante de detonação (considera-se o instante da detonação como sendo o instante de tempo $t=0s$) [Thomas 2001].

A velocidade de propagação das ondas nas rochas é variável segundo uma grande diversidade de fatores, tais como a densidade e as constantes elásticas do meio, que por sua vez são dadas em função de características intrínsecas da rocha, tais como porosidade, temperatura, pressão, entre outros. No entanto, um valor médio de velocidade de propagação tipicamente considerado é de 3000 metros por segundo. Pode-se considerar dessa forma que as profundidades mapeadas são próximas de 6000 metros (dois segundos para o sinal se propagar na ida, mais 2 para refletir até ser registrado pelos sensores).

A imagem sísmica que é obtida por esse processo de captação dos sinais refletidos será melhores quanto menor for o espaçamento lateral entre os sensores localizados na superfície. Valores típicos para o número de sensores estão entre 128 e 1024, situados equidistantes a distâncias que variam de 6,25 a 25 metros [Thomas 2001]. A frequência de amostragem do sinal por cada um dos sensores é o que determina o espaçamento vertical de medição dos sinais, sendo igualmente importante. Equipamentos modernos tipicamente realizam amostragem temporal das ondas refletidas a valores equidistantes de 2 ou 4 milissegundos.

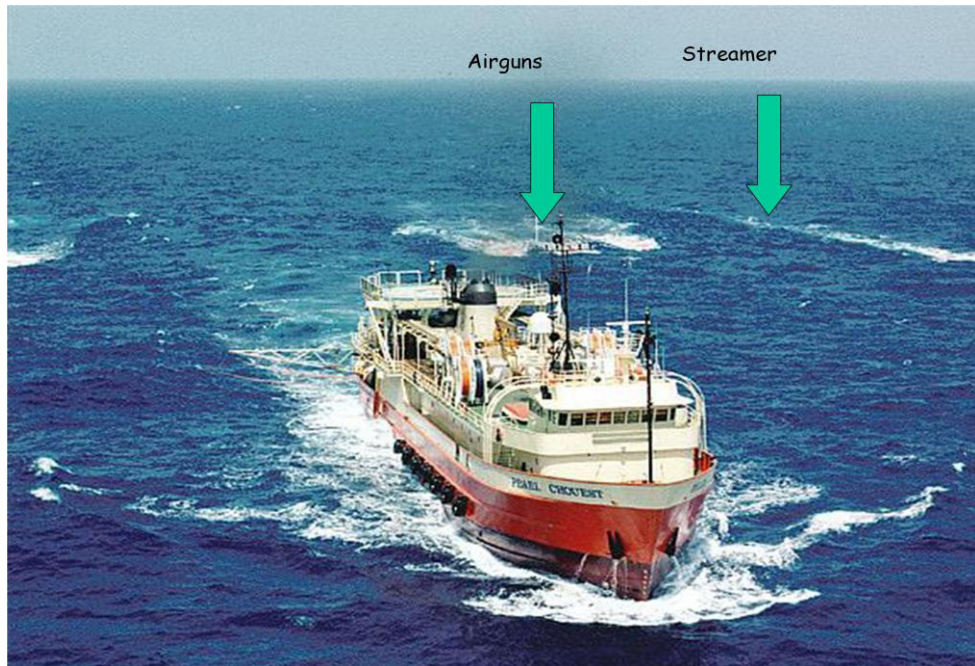


Figura 2.5: Aquisição Marítima [Oilfield].

2.1.4. Processamento

Uma vez realizado aquisição e representação de dados, a fase seguinte consiste em efetuar o processamento dos dados de forma a obter os resultados desejados.

Na sísmica de petróleo, o processamento de dados visa produzir imagens do interior que possuam a maior fidelidade possível, procurando para isso avaliar as distorções geradas pelo processo de aquisição. Nessa etapa, alguns erros próprios ao levantamento sísmico são corrigidos. Além disso, os dados são reorganizados para formarem uma fatia, em relação a imagem do 2D, ou uma grade tridimensional (*voxel*), em relação a imagem de um volume sísmico. Duas das dimensões do conjunto de dados são direções espaciais e estão relacionadas com as posições das fontes e dos receptores. Uma das transformações realizadas nos dados durante o processamento faz com que as posições da fonte e do receptor sejam a mesma. Também em virtude a esta transformação podemos considerar que a terceira dimensão do conjunto de dados é a temporal e que a propagação da onda é feita apenas na direção vertical. Como podemos considerar que a fonte e o receptor estão na mesma posição na superfície, o tempo de cada amostra corresponde ao

tempo que a onda leva para se propagar até uma interface mais o tempo da volta à superfície.

Uma coluna de amostra com as mesmas coordenadas espaciais, variando apenas o tempo, é chamada de traço sísmico. Os máximos e mínimos da função de amplitude sísmica do traço são chamados de eventos sísmicos [Silva, 2004].

A organização das amostras em um dado sísmico é mostrada na Figura 2.6. Do lado esquerdo temos a função de amplitudes sísmicas do traço sísmico, na qual a única dimensão é a temporal (1D). Nesse caso, o traço de amplitudes é representado através de um sinal ondulante. No centro temos uma seção vertical do conjunto de dados formada por um conjunto de traços sísmicos, que é chamada de linha sísmica (2D), com uma dimensão espacial e a outra temporal. No caso dos dados sísmicos 3D (volume sísmico), formados por várias linhas sísmicas, temos duas direções espaciais, que são chamadas de *inline* (direção das linhas sísmicas) e *crossline* (direção perpendicular às linhas sísmicas), além de uma direção temporal. Nesse contexto vale ressaltar que, ao mostrarmos dados em 2D e 3D a forma de representação do sinal de cada um dos traços sísmicos deixa de ser feita a partir de um sinal ondulante, passando a ser feita através de uma escada de cores onde cada cor representa a amplitude da onda em cada ponto.

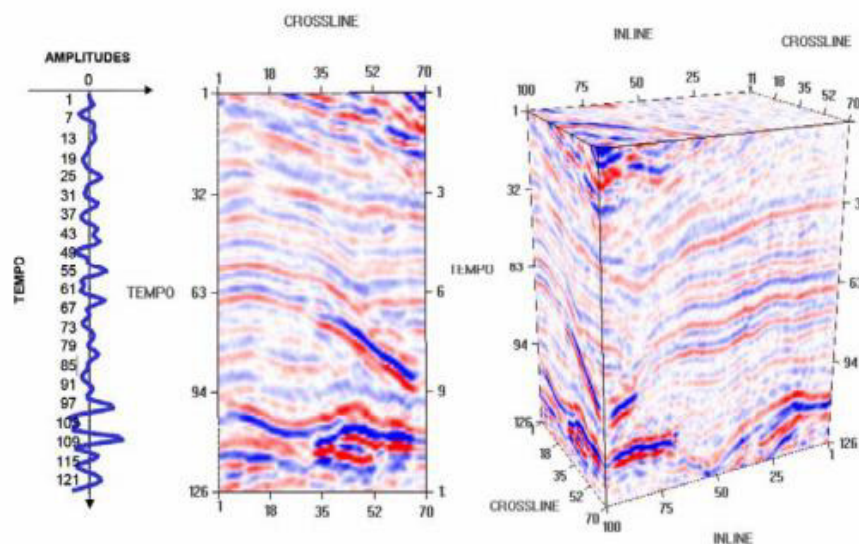


Figura 2.6: Traço sísmico (esquerda), linha sísmica (centro) e volume sísmico (direita). [Kristofer M. Tingdahl & Matthijs de Rooij]

2.1.5. Interpretação

A interpretação sísmica no final da década de 1970 era feita com auxílio de mapas desenhados em papel, dessa forma os especialistas apontavam as áreas de interesse para a perfuração. Os especialistas precisavam ser capazes de criar um modelo tridimensional mental sobre a espessura, constituição e profundidade das camadas de rocha. O trabalho se tornou muito mais simples, uma vez que os especialistas não precisam mais usar sua imaginação para desenharem modelos mentais. Um mapeamento que antes levaria meses para ser realizada, agora pode ser concluído em poucas horas [Petrobrás, 1999].

Nesta etapa de interpretação, o intérprete, em geral um geólogo ou geofísico, analisa os dados sísmicos e tenta criar um modelo que represente a geologia contida na área analisada. A Figura 2.7 mostra um modelo geológico que poderia ser resultante da interpretação de uma linha sísmica. A interpretação sísmica pode ser classificada, de acordo com o foco, em dois tipos: estrutural e estratigráfica. A interpretação estrutural basicamente tenta identificar as camadas geológicas ou, de forma equivalente, as interfaces entre as camadas, bem como as falhas geológicas que recortam as camadas. Na interpretação estratigráfica o foco do trabalho está em entender a maneira como as camadas foram se formando ao longo do tempo.

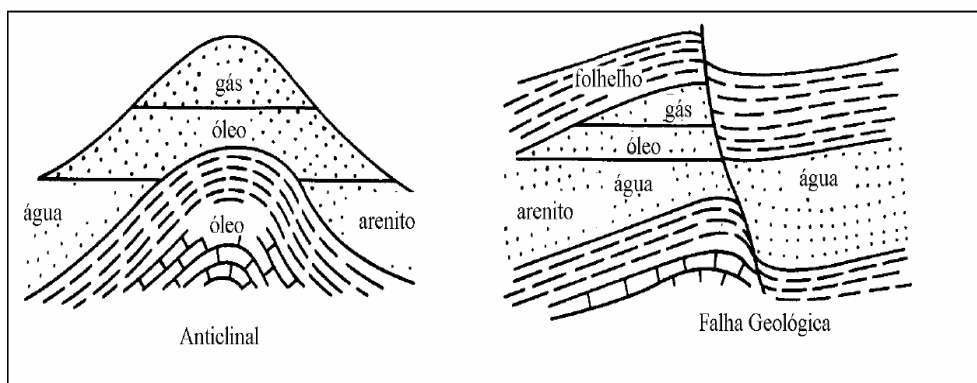


Figura 2.7 – Modelo geológico. Adaptada de [Robison, Treitel 1980].

Um horizonte sísmico pode ser definido como uma série de reflexões contínuas de intensidades similares encontradas através de vizinhanças laterais ao

longo do dado sísmico. Essas reflexões indicam a existência de uma interface entre duas camadas de sedimentos. Sheriff [Sheriff 1991] define um horizonte sísmico como sendo a superfície que separa duas camadas diferentes de rocha, onde tal superfície (mesmo sem ter sido explicitamente identificada) está associada com uma reflexão que se estende por uma grande área. Um horizonte sísmico, portanto, se manifesta em um dado sísmico como uma série de eventos (picos ou vales de amplitudes sísmicas) que aparecem de forma consistente traço a traço. O mapeamento dos horizontes do conjunto de dados é uma das tarefas mais importantes da interpretação sísmica.

Um horizonte sísmico, também chamado de refletor, é uma superfície que separa duas camadas de rocha com propriedades distintas, sendo o limite visível entre determinadas camadas de sedimentos. E as falhas sísmicas são fraturas discretas dos horizontes, através das quais são medidos os deslocamentos da rocha. Não são facilmente identificados nas imagens, mas podem ser percebidas através das descontinuidades lateral do horizonte, como foi mostrado na Figura 2.7.

Na visualização dos resultados da interpretação fornece evidências ou não da presença do acúmulo de hidrocarbonetos. A confiança no resultado da interpretação depende da complexidade do sistema geológico, da qualidade dos dados sísmicos e da experiência do interprete [Rankey & Mitchell, 2003].

2.2. Atributo Sísmico

Atributo Sísmico que descrevem os dados sísmicos, foi popularizado por Taner, no final da década de 1970, define atributo sísmico como “qualquer informação obtida a partir dos dados sísmico seja por medida direta, seja por um raciocínio lógico ou baseado na experiência”[Taner, 2001].

Atributo sísmico surgiu para transformar o processo de interpretação subjetiva e experiência baseada em algo menos tedioso e mais objetivo. Eles tentam imitar experiência do intérprete, identificando automaticamente e correlacionar eventos sísmicos com reais características geológicas [Chopra e MARFURT, 2005]. Hoje, atributos sísmicos são adicionados valores para análise estrutural, estratigráfica e textura, e em fácies e reservatório de hidrocarbonetos de predição de propriedades, quando corretamente utilizado por um usuário experiente (por

exemplo, [Taner, 2001], [Barnes, 2001],[Sheline, 2005],[Chopra e MARFURT, 2005].

Atributos sísmicos permitir a obtenção de informações detalhadas precisas sobre os parâmetros estruturais, estratigráficas e litológicas da perspectiva sísmica [Taner, 2001] como uma solução integrada [Cooper, 2008] para modelagem de reservatório de hidrocarbonetos, identificação e caracterização de incerteza reduzida [Sheline, 2005]. No limite, a questão principal não é os atributos próprios, mas a sua aplicação, que deve ser cuidadosamente escolhido. As características desejadas geológicas devem ser reforçadas deve ser a primeiro claramente definido pelo intérprete.

O principal objetivo dos atributos sísmicos é ajudar o intérprete a gerar um modelo geológico de uma determinada área sísmica. Quanto mais detalhado for o modelo geológico feito pelo intérprete maior será o índice de sucesso na exploração da área em questão.

A análise de características sísmicas é realizada através do emprego de técnicas de reconhecimento de padrões, onde através da combinação adequada de atributos sísmicos, busca-se a identificação de características geológicas dos reservatórios. Neste trabalho foi usada a técnica de Máquinas de Suporte Vetorial (MSV) para classificar as imagens sísmicas, objetivando detectar as falhas sísmicas no dado.

2.2.1. Importância do atributo sísmico na geofísica

Com o crescente interesse nos atributos sísmicos e seu grande número e diversidade torna-se agora necessário catalogá-las em classes diferentes. Muitas propostas foram apresentadas com o objetivo de classificar atributos sísmicos de uma forma firme e rigoroso e intuitivo, com base tanto a entrada quanto o resultado esperado. Infelizmente, novos atributos surgem a cada dia e os algoritmos de atributos bem conhecidos podem ser melhorados, pois algumas vezes eles dão resultados inesperados. As classificações propostas estão constantemente a mudar, dependendo do entendimento do atributo sísmico [Taner, 2000].

O estudo dos atributos sísmicos pode ser utilizado para fazer inúmeras interpretações geológicas e geofísicas tornando de extrema importância para o

geofísico na etapa de interpretação. Dentre os vários atributos existentes, alguns podem ser usados como indicadores de hidrocarbonetos, outros como indicadores de falhas geológicas, ou seja, dependendo do tipo de análise em questão, deve-se escolher o tipo adequado de atributo [Matos, 2004].

Existe uma lista de atributos sísmicos e seus parâmetros Na Figura 2.8, utilizados por Tingdahl e Rooij [Tingdahl & de Rooij, 2005] em seus trabalhos sobre classificação semi-automática de falha em dados sísmicos 3D, que definem diferentes características de falhas e são frequentemente incluídos no esquema de classificação de falha em rede neural artificial.

No contexto desse trabalho, serão aplicados os atributos de similaridade e energia descritos por Tingdahl [Tingdahl & de Rooij, 2005] na possibilidade de detectar a falha.

Attribute	Time gate (ms)	Lateral position	Other settings
Energy	[-32,32]	-	-
Similarity	[-24,24]	(0,1) and (0,-1) 90° rotated	Minimum similarity is output
Steered similarity	[-24,24]	Positions (0,1) and (0,-1) 90° rotated	Minimum similarity is output
Cube similarity	[-24,24]	Full block	Full steering, minimum similarity is output
Cube similarity variance	[-24,24]	Full block	Full steering
Frequency-AFS	[-36,36]	-	-
Frequency-FSF	[-36,36]	-	-
Dip angle	-	-	-
Dip variance	[-32,32]	Stepout 1	-
Most positive curvature	-	Stepout 2	-
Most negative curvature	-	Stepout 2	-
Disc variance	-	Stepout 3	Full steering

Figura 2.8: Atributos utilizados por Tingdahl [Tingdahl & de Rooij, 2005].

2.2.2. Similaridade

A similaridade S é calculada entre dois traços, sendo definido como a distância no hiperespaço entre os vetores dos segmentos, normalizados á soma dos comprimentos dos vetores. Descrito em [Tingdahl & de Rooij, 2005], sendo expresso por:

$$s = 1 - \frac{|v-u|}{|v|+|u|} \quad (2.1)$$

onde,

$$(2.2)$$

$$v = \begin{bmatrix} f(t_1, x_v, y_v) \\ f(t_1 + dt, x_v, y_v) \\ \cdot \\ \cdot \\ \cdot \\ f(t_2 - dt, x_v, y_v) \\ f(t_2, x_v, y_v) \end{bmatrix} \quad u = \begin{bmatrix} f(t_1, x_u, y_u) \\ j(t_1 + dt, x_u, y_u) \\ \cdot \\ \cdot \\ \cdot \\ j(t_2 - dt, x_u, y_u) \\ f(t_2, x_u, y_u) \end{bmatrix}$$

e t é o intervalo de tempo (profundidade) considerado, dt é o intervalo amostral, t_1 e t_2 são os limites do tempo da janela, (x_v, y_v) e (x_u, y_u) são as duas posições de traço que serão comparadas, e f é o valor de amplitude no cubo. Comparado com a coerência tradicional [Bahorich & Farmer, 1995], a similaridade também leva em conta as diferenças de amplitude entre os dois segmentos de traço. Falhas são descontinuidades no dado que dá uma baixa resposta á similaridade.

2.2.3. Energia

A Energia é definida segundo Tingdahl e Rooij [Tingdahl & de Rooij, 2005] como o quadrado da soma da amplitude dentro de uma janela (*gate*) ao redor da amostra de investigação:

$$E = \frac{\sum_{i=0}^{N-1} f(t_0 + i \cdot dt)^2}{N} \quad (2.3)$$

onde, f é a amplitude do traço, t_0 é o limite superior da janela, N é o número de amostras na janela e dt é o intervalo amostral.

2.2.4. Funções Geoestatísticas

A geoestatística surgiu com os trabalhos de [Krige 1951] com dados de concentração de ouro nos quais concluiu que apenas a informação fornecida pela variância seria insuficiente para explicar o fenômeno em estudo. Era preciso

considerar a distância entre as observações. Assim, a geoestatística leva em consideração a localização geográfica e a dependência espacial [CAMARGO 1997].

Segundo [Kourgli e Belhadj-aissa 2004] na análise de textura a primeira e mais importante tarefa é extrair características da textura que mais expressem as informações sobre a distribuição espacial de variações de intensidades em uma imagem. Sendo assim, a aplicação destas funções no reconhecimento de falhas na imagem sísmica é uma alternativa, pois a imagem sísmica possui uma estrutura interessante e provê uma textura rica em informações que estas funções podem capturar.

Os métodos covariograma, semivariograma e correlograma são usados como estimadores geoestatísticos da autocorrelação espacial como ferramenta de continuidade. Essas ferramentas são usadas para investigar a magnitude da correlação entre as amostra e sua similaridade, com a distância.

Na estatística, a textura pode ser descrita em termos de dois componentes principais associados a valores de variabilidade e autocorrelação espacial. Essas medidas descrevem a textura obtida de uma determinada imagem através do grau de associação espacial presente dentro dos elementos espaciais referentes à imagem.

Os métodos de análise de textura baseado em geoestatística como: Covariograma, Semivariograma e Correlograma estão sendo muito utilizado, com sucesso para extrair informações das mais diversas imagens [SILVA, 2004].

Estas funções são aplicadas em diversas áreas, tais como: Ciências da Terra, Sensoriamento Remoto, Processamento de Imagens, etc. Segundo [Srivastava and Parker 1988], nessas áreas de aplicação analisar a continuidade espacial é importante, pois as mesmas trabalham com amostras que apresentam uma grande soma de informações que geralmente se apresentam de forma irregular/ou regularmente espaçadas.

Neste processo para que a extração de característica da imagem seja satisfatória é necessário que a propriedade desses seja descrita de maneira eficiente.

2.2.4.1. Covariograma

O covariograma mede a correlação entre duas variáveis. Em geoestatística, a covariância é calculada como a variância da amostra menos o valor da variograma.

A função de covariância tende a aumentar quando mais próximas as variáveis estiverem, ou seja, quando $h = 0$ e tende a decrescer quando mais distantes, ou próximo do limite, as variáveis estiverem. O covariograma é definido como:

$$C(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} x_i y_i - M_o M_e \quad (2.4)$$

onde M_o é a média dos valores das origens dos vetores,

$$M_o = \frac{1}{N(h)} \sum_{i=1}^{N(h)} x_i \quad (2.5)$$

e M_e é a média dos valores da extremidades dos vetores,

$$M_e = \frac{1}{N(h)} \sum_{i=1}^{N(h)} y_i \quad (2.6)$$

e h é o vetor distância (*lag distance*) entre os valores de origens x_i , e os valores de extremidade, y_i , e $N(h)$ é o número de pares na distancia h . O vetor distância é exemplificado na Figura 2.9.

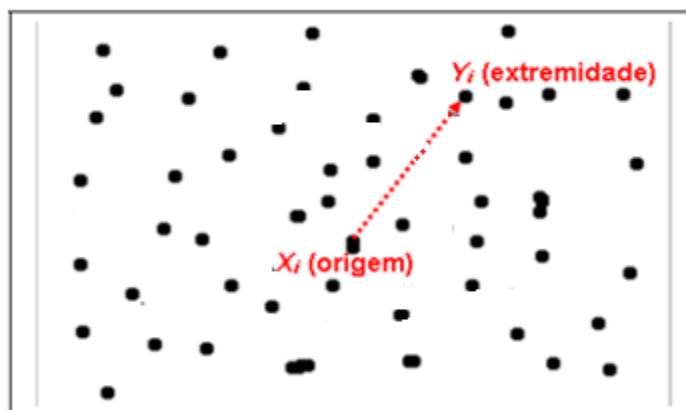


Figura 2.9: Definição do vetor distância.

Os outros parâmetros usados para calcular o covariograma, como incremento do lag (*lag spacing*), tolerância do lag (*lag tolerância*), direção (*direction*), tolerância angular (*angular tolerance*) são ilustrados na Figura 2.10.

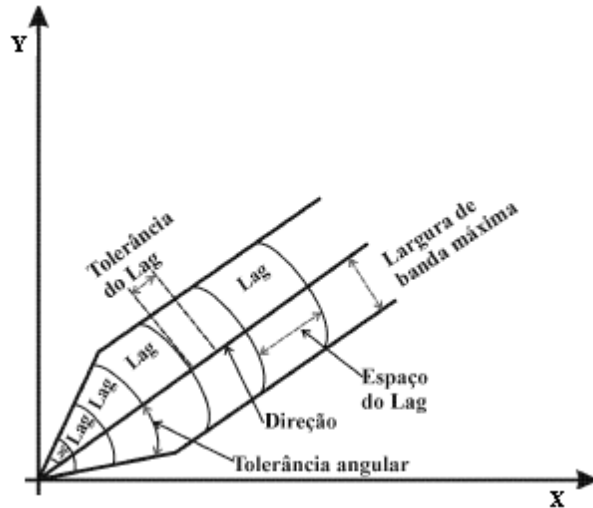


Figura 2.10: Parâmetros utilizados para cálculo das função geoestatísticas.[Aristófanés 2004].

Esses parâmetros são necessários, pois dificilmente um conjunto de pares de pontos da amostra vão estar separados precisamente por uma distância h , logo, são utilizados erros para mais e para menos em função da distância (incremento do lag) e em função da direção (tolerância angular) com o objetivo de alcançar um conjunto mais adequado de pares de pontos que estão relacionados com a distância em questão.

2.3. Reconhecimento de Padrões

Reconhecimento de padrões (RP) é a área de pesquisa que tem por objetivo a classificação de objetos (padrões) em um número de categorias ou classes. As técnicas de RP classificam informações baseada ou em conhecimento pré-definido ou em informações estatísticas extraídas dos padrões.

Os sistemas de RP são formados por um sensor que captam observações a serclassificado. Um mecanismo de extração de característica que discretiza as observações captadas pelo sensor para formar o vetor de características e por último um mecanismo de classificação das amostras observadas que depende das

características que foram extraídas. A classificação pode ser de dois tipos: supervisionada e não-supervisionado.

Na classificação supervisionada, que está sendo utilizada neste trabalho, necessita-se de um supervisor para apresentar previamente o conhecimento do ambiente na apresentação dos dados de entrada e saída, de onde é extraída a representação do conhecimento, com o objetivo das saídas corretas para dados não conhecido anteriormente.

Um dos pontos cruciais de um sistema RP é a escolha de características. As características devem conter a maior parte da informação relevante possível e a dimensão do conjunto de características não deve ser muito grande. Para isso, é importante pré-processar os vetores de características a fim de retirar todas as características desnecessárias. Se duas características são extremamente correlatas, elas são redundantes. Esse tipo de característica pode sobrecarregar o classificador e induzir o processo ao erro.

Após o processo de seleção das características de cada objeto da população, a próxima etapa é atribuir um rótulo a cada vetor de características formando as amostras. Os objetos são rotulados a partir do conhecimento humano. O classificador utiliza as amostras no treinamento. Na fase de treinamento o classificador tenta gerar assinaturas, que melhor distinguem as classes, para cada rótulo pertencente ao conjunto de amostra. Esse processo é importante na fase de conhecimento que fará uso da assinatura para identificar se novas amostras não treinadas fazem parte de uma população específica.

Este trabalho usa Máquina de Vetor de Suporte para realizar o reconhecimento das imagens sísmicas.

2.3.1. Máquina de Vetor de Suporte

As Máquinas de Vetores de Suporte (MVS) constituem uma técnica de aprendizado de máquina, e têm sido amplamente usadas para classificação devido à sua capacidade de generalização, robustez diante de dados de grande dimensão.

As MVS são embasadas pela teoria de aprendizado estatístico (TAE), desenvolvida por [Vapnik, 1998]. Essa teoria estabelece uma série de princípios que devem ser seguidos na obtenção de classificadores com boa generalização,

definidos como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu.

Neste trabalho estamos utilizando o algoritmo supervisionado. Entretanto assume-se que já existe um conjunto de amostra de padrões cujas classes são conhecidas, conforme:

$$S = (x_1, k_1), (x_2, k_2), \dots, (x_m, k_m)$$

onde, X_i representa a i -ésima observação ou padrão. K_i representa a classe, ou rótulo, do padrão acima, e m representa o no de amostras disponíveis. Deve-se produzir um classificador, também denominado modelo, preditor ou hipótese, capaz de prever precisamente o rótulo de novos dados. Esse processo de indução de um classificador a partir de uma amostra é denominado treinamento [Lorena & Carvalho, 2002].

Em uma tarefa de classificação normalmente envolve treinamento e teste de dados que consiste em alguns exemplos de dados. Cada exemplo no conjunto de treinamento contém um “valor objetivo” (rótulos de classe) e vários “atributos” (características), e frequentemente, a variável de interesse é definida da seguinte forma:

$$y_i = \begin{cases} 1, & \text{se o indivíduo pertence ao grupo 1} \\ 0, & \text{se o indivíduo pertence ao grupo 2} \end{cases} \quad (2.7)$$

onde o principal interesse é obter $Prob(y = 1/x) = f(x, \beta)$, e na prática, $f(x, \beta)$ é sempre uma função desconhecida.

2.3.2. Considerações sobre a Escolha do Classificador

O objetivo do MVS é produzir um classificador que apresente um bom desempenho junto a amostras não-observadas durante o treinamento, isto é, consiga generalizar. Um classificador é uma máquina que fornece classificação. No exemplo da Figura 2.11, observamos a existência de vários classificadores lineares possíveis que podem separar as amostras disponíveis sem nenhum erro, mas há somente um que maximiza a distância entre o classificador e a amostra mais próxima de cada classe. Dessa forma esse classificador linear é denominado

hiperplano ótimo, pois, espera-se que este hiperplano generalize melhor que os demais, quando amostras não utilizada durante a fase de treinamento devem ser classificadas [Lima, 2004]. Da mesma forma podemos observar na Figura 2.11 que o hiperplano ótimo criado tem outros hiperplanos como suporte (linhas tracejadas) denominado vetores-suporte e que passam por pontos das classes.

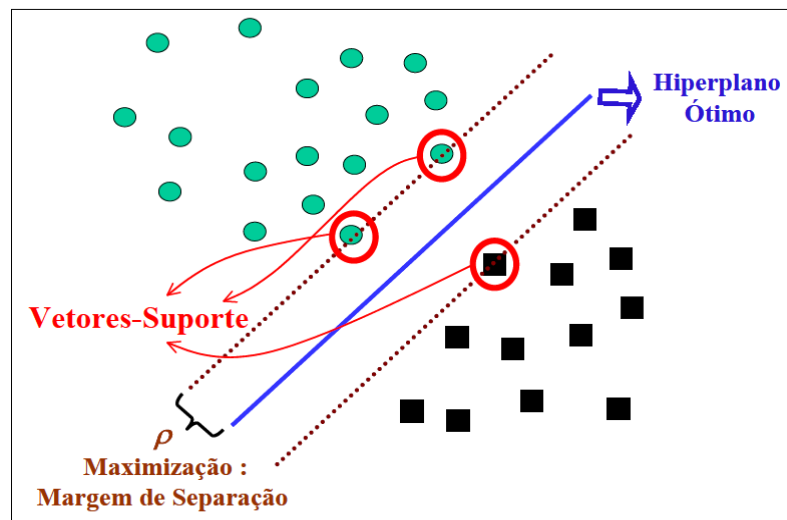


Figura 2.11: Hiperplano ótimo.[Lima, 2004]

O MVS para encontrar a solução ótima utiliza o princípio de minimização do risco estrutural e baseia-se no fato de que a taxa de erro nos dados de teste (taxa de erro de generalização) é limitada pela soma da taxa de erro de treinamento e por um termo que depende da dimensão de Vapnik-Chervonenkis (dimensão VC)³

Por exemplo, dado um conjunto de treinamento de pares de rótulo (x_i, y_i) , $i = 1 \dots l$ onde $x_i \in R^n$ e $y_i \in \{1, -1\}$. A classificação da margem de separação ao máximo tem como finalidade separar as duas classes por um hiperplano tal que a distância dos vetores suporte seja maximizada. Cada ponto x no espaço de entrada é mapeado para um ponto $z = \Phi(x)$ de alta dimensionalidade, denominado espaço de característica, onde os dados são linearmente separados por um hiperplano.

Esses dados podem ser divididos em três categorias os chamados linearmente separáveis, não linearmente separáveis e não separáveis. Dessa forma nós podemos escrever a nossa função de decisão usando uma função de *kernel* $K(x, y)$, por ser mais simples que o do mapeamento Φ , e que é definida pela função:

$$f(x) = \sum_{i=1}^l \alpha_i y_i K(x, x_i) + b \quad (2.8)$$

onde $k(x, xi) = \Phi(x) \cdot \Phi(xi)$, e os coeficientes $aieb$, são soluções das equações:

$$\min_{w,b,\xi} \frac{1}{2} w^T \cdot w + C \sum_{i=1}^l \xi_i \quad (2.9)$$

sujeito a:

$$y_i [w^T \cdot \phi(x_i) + b] \geq 1 - \xi_i \quad (2.10)$$

onde w é um vetor de pontos perpendicular no hiperplano de separação, $C > 0$ é um parâmetro escolhido pelo usuário, correspondendo a penalidade do erro e os ξ_i 's são variáveis de folga que penalizam os erros de treinamento.

As MVS possuem diferentes *kernels* que são utilizados para resolver problemas de espaços não lineares, sendo os mais utilizados os tipos Linear, Polinomial (que manipula uma função polinomial cujo grau pode ser definido durante os treinamentos), Sigmoidal (permite que a SVM tenha o comportamento de uma rede MLP4) e Gaussiano (corresponde a um espaço de características de dimensão infinita; a utilização desse tipo de função permite que a MVS se comporte com características de uma rede RBF5). Temos as quatro funções básicas de *kernel*:

- Linear: $K(x, y) = x^T y$;
- Polinomial: $K(x, y) = (\gamma x^T y + r)^d, \gamma > 0$;
- Sigmoidal : $K(x, y) = \tanh(\gamma x^T y + r)$;
- Função básica radial (RBF): $k(x, y) = e^{-\gamma \|x-y\|^2}$.

As diferentes definições do *kernel* e seus respectivos parâmetros provocam alterações nos resultados fornecidos por uma SVM. Utilizamos nesse trabalho o *kernel RBF* com o parâmetro $\gamma > 0$ definido pelo usuário.

2.3.3. Validação de Resultados

O algoritmo de AM supervisionado é normalmente realizado por meio da análise do desempenho do preditor gerado na classificação de novos dados, não apresentados previamente em seu treinamento [Lorena & Carvalho, 2002].

Utilizaremos nesse trabalho a Sensibilidade (s), Especificidade (e) e a Acurácia(a), por serem métricas muito utilizadas, para validar os resultados.

A sensibilidade é a probabilidade de positivos verdadeiros, ou seja, a medida da capacidade do método de decisão de prever a condição Falha para aqueles casos que realmente a apresentam, segundo a definição da equação.

$$s = \frac{VP}{VP+FN} \quad (2.11)$$

onde, VP são pontos definidos na imagem como Falha e que foram classificados como tal e FN o número de falso negativo, no qual pontos definidos como Falha são classificados como de Não-Falha classificado erroneamente.

A especificidade (e) é a probabilidade de negativos verdadeiros, ou seja, a medida da capacidade do método de decisão de apontar ausência da condição para aqueles casos que realmente não a têm, segundo a definição da equação:

$$e = \frac{VN}{VN+FP} \quad (2.12)$$

onde VN é o número de verdadeiros negativos, Não-Falha classificado como Não-falha, e FP é o número de falsos positivos, Não-Falha classificado como Falha.

A acurácia (a) é a probabilidade de predições corretas, e é dada pela equação abaixo:

$$a = \frac{VP+VN}{VP+VN+FP+FN} \quad (2.13)$$

3. MATERIAIS E MÉTODOS

Este trabalho visa estudar o problema de classificação de falhas em volumes de dados sísmico e utilizando aprendizado de máquina, visando, assim, auxiliar o intérprete, na etapa de interpretação da imagem sísmica. Para esse fim seguimos alguns passos:

1. Aquisição e tratamento das imagens sísmicas;
2. Cálculo dos atributos sísmico de falha: Energia Vertical e Similaridade Vertical;
3. Cálculo do atributo Covariograma como descritores de característica.
4. Classificação usando SVM.

3.1. Aquisição e tratamento das imagens sísmicas

Nesse trabalho foi necessário fazer o tratamento da imagem. Sendo os dados armazenados no formato *begin-endian*, byte menos significativo com o maior endereço. Os dados do volume foram convertidos para *little-endian*, byte menos significativo no menor endereço. Tratamento das imagens sísmicas.

Nesse contexto as fatias que são obtidos do volume. Todas as imagens estão na resolução 256 x 256 pixels, em tons de cinza de 8bits. Os dados de amplitude, que são valores entre -1699.143604 e +1769.342529, são quantizadas para 8bits gerando imagens conforme a Figura 3.1.

3.2. Cálculo dos atributos sísmico de falha

Na literatura existem vários atributos sísmicos de identificação de falhas, esses atributos identificam características específicas das imagens. Entre os existentes utilizamos para esse propósito dois atributos, energia e similaridade. Para cada atributo, foi calculado ao longo do traço sísmico, na vertical da imagem. Os atributos calculados foram: Similaridade Vertical (SV), Energia Vertical (EV).

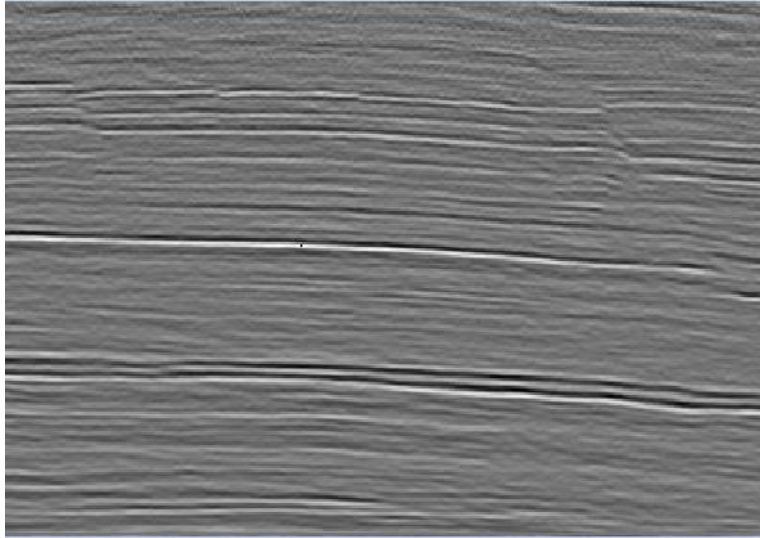


Figura 3.1: Fatia gerada a partir de dados da amplitude.

3.2.1. Cálculo da Similaridade Vertical (SV)

O atributo SV, foi calculado usando a Equação 2.1 e 2.2 que se encontra na Seção 2.2.2. Esses atributos são calculados fatia por fatia, sempre definindo o vetor onde esta contida o pixel de referencia, para o qual será calculado o atributo, de v e o vetor adjacente de u , ambos com o mesmo timegate (do inglês, janela). Posteriormente o cálculo realizado, o resultado é armazenado na posição equivalente do pixel de referência. Nesse trabalho definimos timegate's de 7 e 15 para nossas imagens, sendo possível maior extração de características. A Figura 3.2 ilustra o cálculo da SV.

Para calcular a SV, por exemplo, selecionamos o elemento de valor 103, pixel circulado em vermelho, onde foi calculado $|v - u|$, fazendo $|(8 - 9), (1 - 7), (103 - 109), (6 - 7), (4 - 4)|$, dividindo pelo $|v| + |u|$ e subtraindo de 1. O valor calculado é armazenado na posição do elemento 103. Como foi feito o cálculo do atributo de EV, a janela vai sempre percorrendo a imagem na vertical para o cálculo da SV.

Temos nas Figuras 3.3 exemplo de imagem gerada a partir do cálculo do atributo SV.

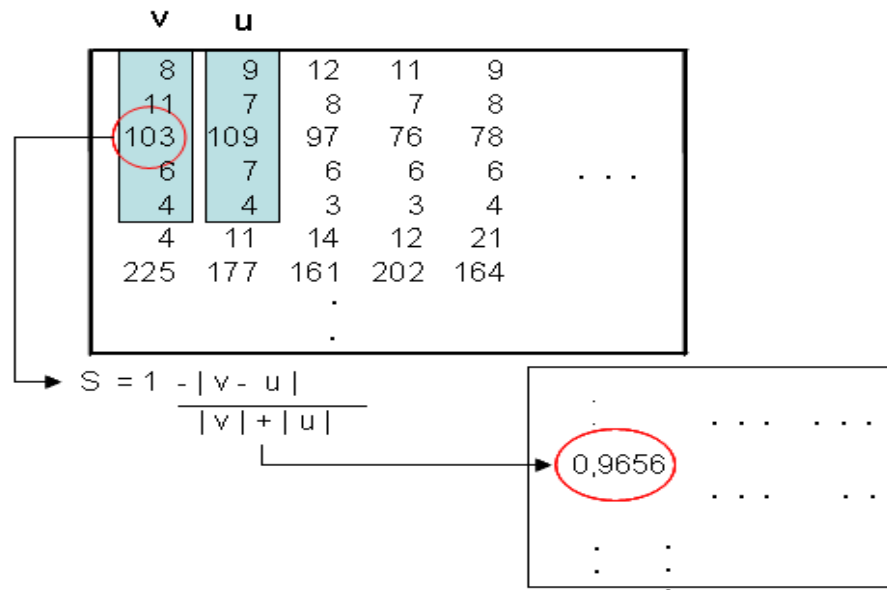


Figura 3.2: Cálculo da Similaridade Vertical.

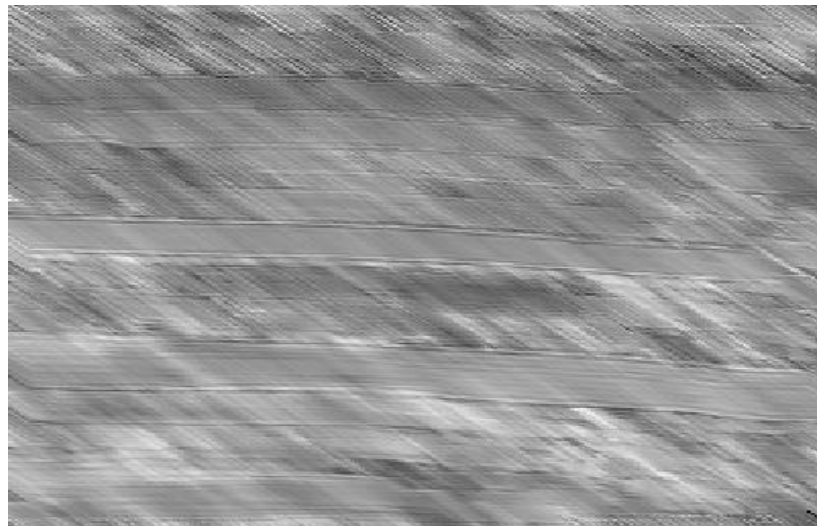


Figura 3.3: Similaridade Vertical

3.2.2. Cálculo do Atributo Energia Vertical (EV)

O atributo EV, foi calculado usando a Equação 2.3 que se encontra na Seção 2.2.3. Para calcular a Energia, por exemplo, selecionamos o elemento de valor 103, pixel circulado em vermelho. O cálculo é efetuado para esse pixel calculando o quadrado da soma da amplitude, $(8^2 + 11^2 + 103^2 + 5^2 + 4^2)$, dentro do *gate* ao redor da amostra de investigação e dividimos por 5, que representa a

quantidade de elementos no *gate*. Essa janela vai sempre percorrendo a imagem na vertical. A Figura 3.4 ilustra o cálculo da EV.

Temos nas Figuras 3.5 exemplo de imagem gerada a partir do calculo do atributo EV.

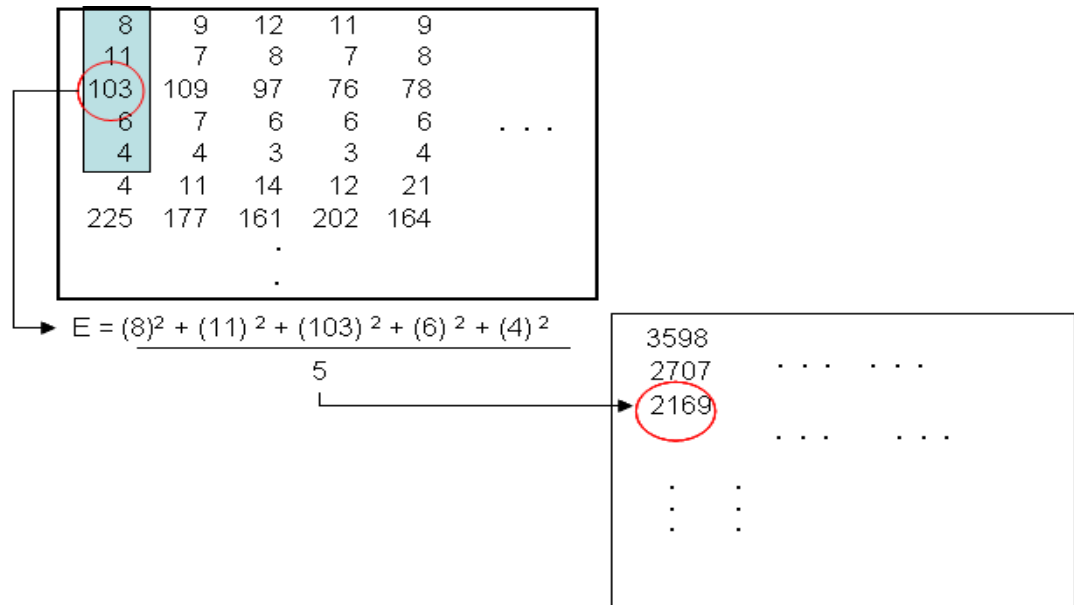


Figura 3.4: Cálculo da Energia Vertical.

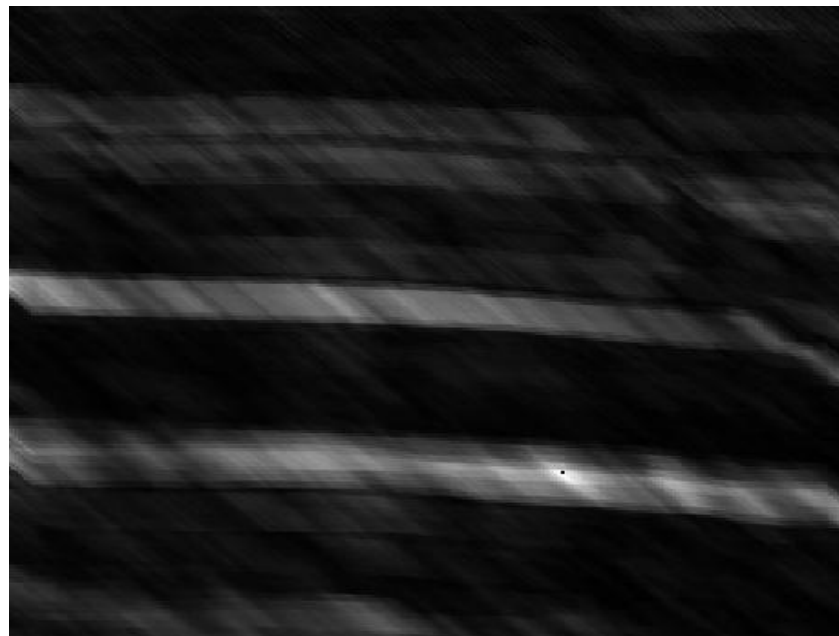


Figura 3.5: Energia Vertical

3.2.3. Cálculo do atributo Covariograma.

Este trabalho utiliza a função geoestatística Covariograma para extrair características da imagem sísmica.

3.2.3.1. Extração de Característica

Reconhecimento de padrões em imagens é uma área de grande interesse no mundo científico. Os chamados métodos de extração de características possuem as habilidades de extrair características das imagens e também de reduzir a dimensionalidade dos dados gerando assim o chamado vetor de características.

Extração de características envolve simplificar o conjunto de dados requeridos para descrever um grande conjunto com mais precisão. Quando executando a análise dados complexos um dos maiores problemas provêm do número de valores envolvidos. Analisar uma gama muito grande de variações numéricas requer grande quantidade de memória e poder de processamento ou um algoritmo de classificação estatística que sobrecarrega a amostra de treinamento e faz pouca generalização para novas amostras (gerando muitas classes, ou classes desnecessárias). Extração de características é um termo genérico para métodos para construção de combinações de valores para contornar esses problemas e continuar representando os dados com certa precisão.

Os parâmetros usados pela função geoestatística para extração das características e, cada amostra foram às direções de 0°, 45°, 90° e 135° com tolerância angular de 22,5° e incremento de *lag* (distancia) igual a 1, 2 e 3 correspondendo a 15, 8 e 5 *lags* e tolerância de cada distancia de *lag* igual a 0,5, 1,0 e 1,5 respectivamente. Esses parâmetros não foram escolhidos ao acaso. As direções adotadas são as mais utilizadas na literatura para análise de imagens, já para escolher a tolerância de *lag* segundo [ISAACS e SRIVASTAVA 1990] a escolha mais comum é adotar a metade do incremento de *lag*.

Na Figura 3.6, temos a ilustração da maneira como as distâncias e direções do *lag* são utilizadas. Supondo que essa imagem seja de 15 x 15 pixels, que é o tamanho da janela utilizada, o máximo de lags que teremos em qualquer das quatro direções para $d = 1$ será 15, $d = 2$ será 8 e $d = 3$ será 5 justificado assim, a presença desses valores nos parâmetros.

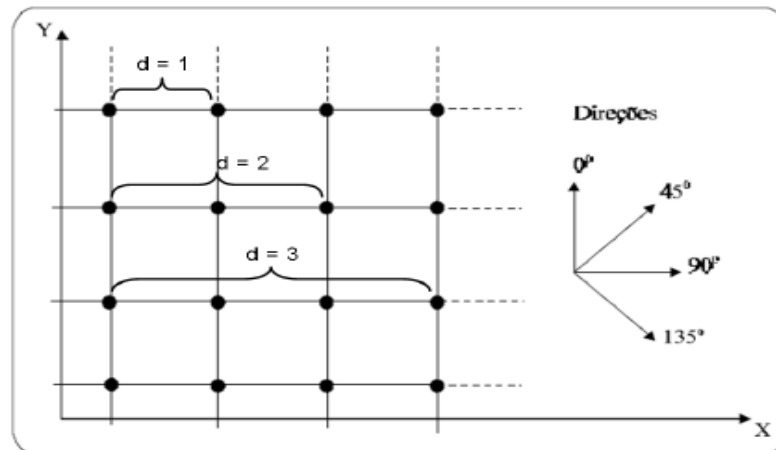


Figura 3.6: Distância e direções dos lags. [CAMARGO 1998].

3.2.3.2. Vetor de Características

Para criação do Vetor de Características (VC), que representa a assinatura da amostra, foram extraídas 112 características por amostra correspondentes as 4 direções de 28 lags ($15 + 8 + 5$) para função geoestatística covariograma. Antes de realizar a seleção de características as mesmas passam por um processo de normalização para uma faixa de valores comum como -1 á +1. Esse mecanismo facilita o classificador a convergir com maior facilidade na etapa de treinamento. Além de padronizar a distribuição de valores das variáveis, as quais podem assumir diferentes domínios.

Um VC foi criado para função geoestatística covariogramatem como o objetivo de analisar a eficiência da área de interesse do dado sísmico. Os testes e resultados serão apresentados no Capítulo 4.

3.4. Classificação usando MVS.

A etapa final consiste em classificar cada objeto na imagem sísmica como Falhas e não-Falhas utilizando técnicas de reconhecimento de padrões (seção 2.4) de acordo com a análise de textura obtida pela função geoestatística covariograma (seção 2.3.1). A Figura 3.7 ilustra o fluxo de atividades realizadas durante a fase de reconhecimento de padrões.

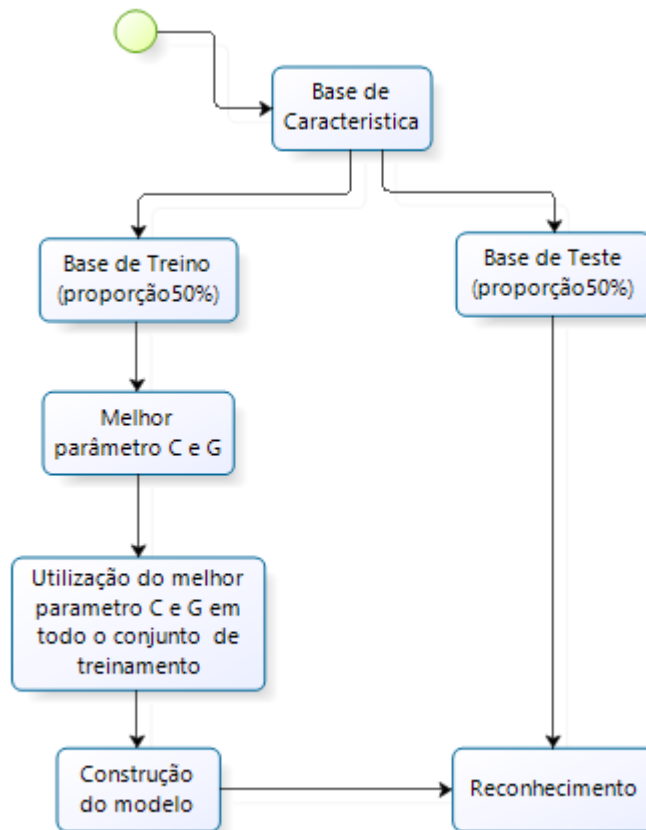


Figura 3.7: Fluxo de atividade realizado durante a etapa de classificação desde a base de características originais, estimação de parâmetros e reconhecimento de padrão.

Nesta etapa fizemos utilização da biblioteca LIBSVM [Chang & Lin, 2001], que tem por finalidade ajudar usuários a usar SVM para fazer o treinamento. Um detalhe importante sobre LIBSVM é que tenta auxiliar as pessoas que nunca treinaram um SVM antes. Para usar o SVM há dois scripts python - grid.py e

easy.py- no “diretorio de ferramentas”. Para classificar o dado sísmico usando SVM, seguimos a seguinte ordem:

- Geração do arquivo de entrada no formato LIBSVM;
- Utilização do $Kernelk(x, y) = e^{-\gamma||x-y||^2}$;
- Utilizamos a base modelo e transformamos para o intervalo de 0 a 1, usando svm-scale;
- Dividimos o esse arquivo em arquivo de teste e arquivo de treino na proporção de 50%, usamos para essa finalidade o script subset.py;
- Utilizamos o script grid.py que implementa a busca automatizada (search grid)para encontrar o melhor parâmetro C e (γ);
- Utilização do melhor parâmetro C e γ em todo o conjunto de treinamento;
- Construção do modelo;
- Classificação do arquivo de teste;
- Calculo das métricas;

Os dados de EV submetidos á Classificação, foi normalizados para o intervalo [0,1] como em:

$$X_{norm} = \frac{x-x_{min}}{x_{max}-x_{min}} \quad (2.16)$$

onde, X_{norm} é o valor normalizado correspondente ao valor original, e x_{min} e x_{max} são os valores mínimo e máximo entre todos os valores. Para a SV não foi necessário fazer a normalização, visto que o calculo da SV já esta entre valores do intervalo [0,1].

Foram definidas duas classes para essa finalidade, Falha (+1) e Não-Falha (-1). Geramos arquivos com 1 atributo de Covariograma, com 1 atributo de SV e com 1 atributo de EV, em seguida foi gerado com 2 atributo de EV e SV e com associação dos três .

Usamos o *kernel* onde há a necessidade de configurar dois parâmetros existentes o C e γ . Não se sabe anteriormente qual C e γ é melhor para o problema, sendo assim, algum tipo de modelo de seleção (parâmetros de busca) deve ser feito, com a finalidade de identificar o melhor C e γ de forma que o classificador possa prever dados desconhecidos com precisão.

Separamos os dados de treinamento em duas partes. Uma é considerada desconhecida no treinamento do classificador. Então a precisão da predição, neste conjunto, pode refletir mais precisamente no desempenho da classificação de dados desconhecidos.

Nessa etapa encerramos com os treinamentos e testes do MVS, configurada com os parâmetros C e γ para cada conjunto de amostras submetidas ao processo.

Na etapa seguinte trataremos dos resultados alcançados com o uso desta metodologia.

4. RESULTADOS

Este capítulo apresenta os resultados obtidos pela metodologia proposta, baseada no método geoestatístico covariograma para classificação de falhas em dados sísmica. Aplicamos esses atributos para 50 fatias, dentro de um volume sísmico com numerosas falhas, onde de cada fatia foi extraído 25 pontos de falhas e também 25 pontos de não-falhas, no total de 2500 amostras. Com o objetivo da classificação de falhas, teve muito eficiência o método de classificação com uso do SVM.

Seguindo o fluxo da metodologia proposta na Seção 2.3.1, o passo seguinte é a etapa de classificação e validação de resultados. Os resultados obtidos pelo classificador SVM estão relacionados na Tabela 4.1.

Tabela 4.1: Resultados da acurácia para 2500 amostra calculada para cada atributos.

	Janela				
		7		15	
Atributo	(%)	Treino	teste	treino	teste
EV	50/50	66,08%	66,88%	68,8%	70,08%
SV	50/50	59,76%	62,24%	62,48%	61,44%
Covariograma	50/50	96,64%	71,0933%	99,04%	76,64%

Dos resultados apresentado na Tabela 4.1, observamos que os atributos isoladamente não apresentaram bons resultados, visto que o Covariograma, que foi o melhor, obteve apenas 76,64% de acurácia. Percebemos também que, no geral, os melhores resultados, para esses testes, foram alcançados com a janela (*time gate*) de 15.

Tabela 4.2: Resultados da acurácia para 2500 amostras calculadas para associação dos atributos.

	Janela				
		7		15	
Atributo	(%)	Treino	teste	treino	teste
EV /SV	50/50	72,72%	70,56%	77,92%	76,24%
EV/SV/ Covariograma	50/50	85,66%	83,52%	94,35%	92,65%

Avaliando os resultados da Tabela 4.2, percebemos que a associação dos 3 três atributos alcançou bons resultados. Com a janela de tamanho 15 e para 50% das amostra no treinamento e 50% nos testes, a acurácia(a) atingiu 92.65%na validação e 94.35% para treinamento.

A partir das Tabelas 4.1 e 4.2, separamos os melhores resultados para calcular a sensibilidade (s) e especialidade (e) gerando a Tabela 4.3:

Tabela 4.3: Melhor resultados obtidos para a classificação.

	Janela de tamanho 15						
		Sensibilidade (%)		Especialidade (%)		Acurácia (%)	
Atributo	Amostras	treino	teste	treino	teste	treino	teste
EV	50/50	60,38	61,20	77,08	80,32	68,8	70,08
SV	50/50	49,22	51,34	67,35	71,57	62,48	61,44
Covariograma	50/50	86,98	87,10	90,75	83,92	99,04	76,64
EV /SV	50/50	71,87	74,23	78,77	84,43	77,92	76,24
EV/SV/ Covariograma	50/50	90,93	91,10	95,76	93,92	92,89	92,23

O melhor resultado obtido alcança sensibilidade de 91.10%, especificidade de 93.92% e acerto de 92.23% para a configuração da função geoestatística. O gráfico da Figura 3.8 mostra os resultados dos testes apresentados na Tabela 4.3, comprovando a eficiência da utilização dos atributos.

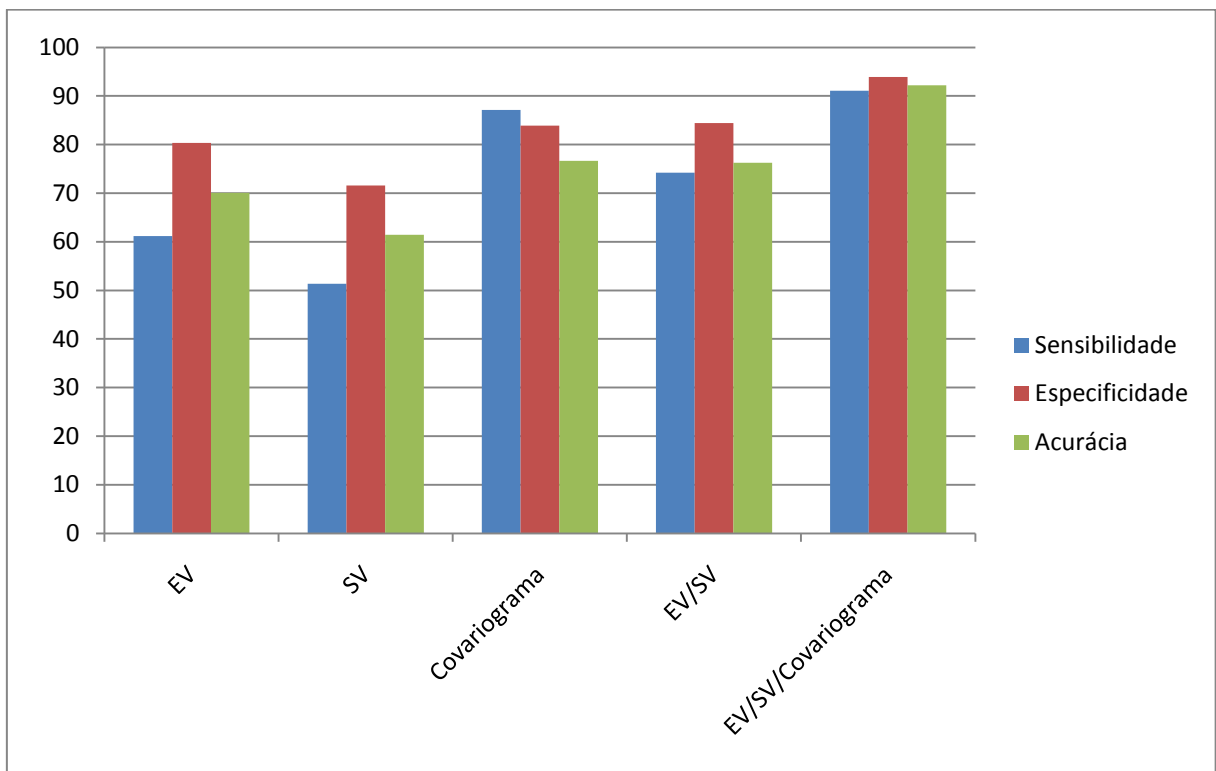


Figura 3.8: Resultados obtidos nos testes.

5. CONCLUSÃO

O presente trabalho apresentou a viabilidade da utilização de técnicas de processamento de imagens, funções geoestatísticas e Máquina de Vetores de Suporte para classificação de falhas em dados volumétricos.

O petróleo é a principal fonte de energia utilizada no mundo, mas também exige um custo bastante elevado para sua obtenção. A perfuração de um poço de petróleo pode atingir, em uma estimativa, até 85% do custo total da exploração. Portanto, é necessário ter o máximo de informações possíveis das condições geológicas da área antes de iniciar a perfuração de um poço. Uma das etapas de obtenção dessa fonte energética é a prospecção que inicia pelos estudos geológicos de observação até a obtenção de dados sísmicos e a perfuração do poço pioneiro. A sísmica hoje é uma das ferramentas mais importantes na prospecção do petróleo, mas os dados sísmicos coletados necessitam de tratamentos através de programas computacionais para poderem ser interpretados pelos geólogos e geofísicos.

Um ponto que merece atenção é a configuração da janela. Uma janela muito grande para uma região pequena, ou vice-versa, pode dificultar a classificação da falha e resultar em conclusões inconsistentes, pois serão calculados atributos em regiões fora da área de interesse fazendo com que os pixels dentro da janela apresentem características semelhantes, ou seja, pontos classificados como falha pode assumir características de não-falha após o cálculo dos atributos.

Foi apresentado, neste trabalho, um método sobre uma metodologia para classificação de falhas em dados sísmicos. Como visto a escolha correta dos atributos, para o objetivo que se deseja obter, é de extrema importância.

Fundamentado nos resultados de especificidade igual a 93.92%, sensibilidade igual 91.10% e acurácia igual a 92.23%, o método apresentado aqui mostra que falhas sísmicas podem ser detectadas, utilizando a função geoestatística covariograma aplicados numa máquina de suporte vetorial. Contudo, é necessário realizar testes com uma quantidade maior de volumes sísmicos, de dados reais, e com diferentes características de falhas, ou seja, imagens mais complexas para que seja possível comprovar a eficiência da técnica proposta.

Na indústria de petróleo, a interpretação sísmica visa identificar estruturas onde o óleo possa estar aprisionado em um reservatório, logo necessita de um conjunto de ferramentas que dê suporte com bastante precisão. Baseado nisso, acreditamos que apesar dos resultados obtidos terem sido bons, é possível melhorá-los, visto que foi utilizada apenas uma função geoestatística. Propomos então como trabalhos futuros:

- Pesquisar outras funções geoestatísticas a serem utilizadas na extração das características. Como por exemplos, a função semivariograma, correlograma e semiridrograma. E fazendo combinações com todas elas.
- O classificador SVM que desempenhou um papel fundamental na classificação, podendo ser trocado por algum outro classificador de maneira a avaliar o desempenho desse novo classificador.

REFERÊNCIAS

BRAZ JUNIOR. Classificação de Regiões de Mamografias em Massa e Não Massa usando Estatística Espacial e Máquina de Vetores de Suporte. Dissertação (Mestrado), Universidade Federal do Maranhão, Engenharia Elétrica, 2008.

CAMARGO, E. C. G. Desenvolvimento, implementação e teste de procedimentos geoestatístico (KRIGEAGEM) no sistema de processamento de informações georeferenciadas (APRING). Dissertação (Mestrado), Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1997.

CARVALHO, M. M. Determinação de Malhas de Falhas em Dados Sísmicos por Aprendizado Competitivo. Tese (Doutorado), Programa de Pós-Graduação em Informática, PUC-RJ, 2008.

CHEVRON, CORPORATION. Disponível em: <http://ep.chevrontexaco.com.br>.

DALLYSON, S. A. J. Detecção de Falhas de Dados Sísmico usando Máquinas de Vetor de Suporte. Trabalho de Conclusão de Curso (Graduação em Bacharelado Em Ciência da Computação) - Universidade Federal do Maranhão, 2007.

DORN, G. A. Modern 3d seismic interpretation. The Leading Edge, v.17, p. 1262-1273, 1998.

GERHARDT, ANDRÉ. Aspectos da Visualização Volumétrica de Dados Sísmicos. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 1998.

KOURGLI, ASSIA; AICHOUCHE, BELHADJ-AISSA. Characterising textural primitives using variography. IMVIP - Irish Machine Vision and Image Processing Conference. Belfast, Ireland, p. 165–175, 2000.

LINES, L.R.; NEWRICK, R.T. Fundamentals of geophysical interpretation. SEG, Society of Exploration Geophysicists. Department of Geoscience, University of Calgary, Calgary, AB T2N 1N4, Canada, 2004.

LIMA. Uma Abordagem Unificada Empregando Máquinas de Vetores-Suporte. Universidade Estadual de Campinas, 2004

MACHADO, MARCOS. Segmentação de Dados Sísmica Via Hyper-stack para Visualização. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro – RJ, 2000.

MATOS, MARCÍLIO CASTRO. Reconhecimento de padrões sísmicos utilizando análises tempo-frequência. Tese doutorado, Pontifícia Universidade Católica do Rio de Janeiro, RJ, 2004.

MATOS, MARCÍLIO CASTRO. Reconhecimento de padrões sísmicos utilizando análises tempo-frequência. Tese doutorado, Pontifícia Universidade Católica do Rio de Janeiro, RJ, 2004.

MORAES. Mapeamento automático de horizontes e falhas em dados sísmicos 3D baseado no algoritmo de gás neural evolutivo. Tese (Doutorado), PUC – RJ, 2008.

OILFIELD, SCHLUMBERGER GLOSSARY. Disponível em:
<http://www.glossary.oilfield.slb.com>.

PEDERSEN, S.; SCHLUMBERGER TECHNOLOGY CORPORATION. Image feature extraction. United States Patent, 7203342, April, 10, 2007.

RANNEY, E. C.; MITCHELL, J. C. Interpreter's Corner– That's why it's called interpretation: Impact of horizon uncertainty on seismic attribute analysis. The Leading Edge, Vol. 22, p. 820, 2003.

SILVA, PEDRO MARIO CRUZ. Visualização Volumétrica de horizontes em Dados Sísmico 3D. Tese doutorado, Pontificia Universidade Católica do Rio de Janeiro, 2004.

SRIVASTAVA, M. R.; PARKER, H. M. Robust measures of spatial continuity. *Geoestatics Proceedings of the Third Int. Geostatistics, Congress, Avigon, France 1*, 295–308, 1988.

TANER, M. T. *Seismic Attributes*. Canadian Society of Exploration Geophysicists. Sept.2000.

TINGDAHL, K. M.; DE ROOIJ, M. Semi-automatic detection of faults in 3D seismic data. *Geophysical Prospecting*, vol. 53, p. 533–542, 2005.

THOMAS, J. E. *Fundamentos de Engenharia de Petróleo*. Ed. Sinopse, 2^o Edição, 2001.